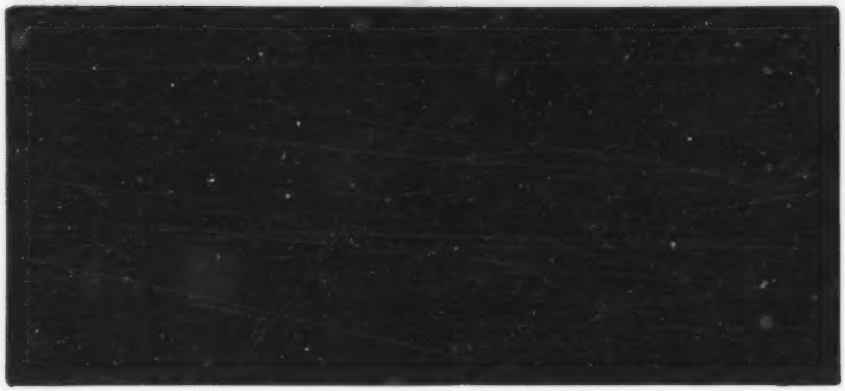
Working Paper

Document de travail



Fiscal Policy and
Economic Analysis
Branch

Department of Finance

Direction de la
politique fiscale
et de l'analyse économique

Ministère des Finances

101-06191/

Économies d'échelle: Un survol des techniques d'estimation et des problèmes méthodologiques

par

BENOÎT ROBIDOUX

Document de travail No. 87-1

Division des études économiques et de l'analyse de politiques
Direction de la politique fiscale et de l'analyse économique
Ministère des Finances, Canada

L'auteur tient particulièrement à remercier John Lester et François Delorme dont les nombreux commentaires ont permis d'apporter des améliorations substantielles à ce texte. Le document a également bénéficié des commentaires de Melvyn Fuss, Pierre Mohnen, Raynald Létourneau, Paul-Henri Lapointe, Kevin Lynch, Munir Sheikh et Peter Gusen. L'auteur reste, toutefois, seul responsable du texte tel que rédigé.

Les documents de travail sont publiés, dans la langue utilisée par les auteurs, afin de susciter des commentaires et d'assurer une diffusion plus large des études analytiques préparées au ministère des Finances. Ces documents n'ont fait l'objet que d'une évaluation sommaire et leur contenu représente le point de vue de leurs auteurs. Les lecteurs sont invités à envoyer leurs commentaires directement aux auteurs.

RÉSUMÉ

L'objectif de ce document est de recenser les diverses méthodes qui ont été proposées dans la littérature pour estimer les économies d'échelle et d'identifier les problèmes auxquels ces méthodes sont confrontées. Les méthodes d'ingénierie, probabiliste, économétrique et les techniques ad hoc basées sur la survie ou la distribution des établissements sont autant d'approches qui ont été proposées dans la littérature d'organisation industrielle et qui sont présentées dans le présent document. De plus, ce survol discute de ces méthodes et de leurs limites à l'intérieur d'un cadre microéconomique.

Même si toutes ces méthodes d'estimation sont présentées, nous insistons particulièrement sur l'approche économétrique. Trois types de fonction de coût sont présentés: (i) les fonctions flexibles non-homogènes et non-homothétiques; (ii) les fonctions homothétiques mais non-homogènes; et (iii) les fonctions homothétiques et homogènes. Les conséquences et les avantages résultant de l'hypothèse d'homothéticité du processus de production sont également présentés. Finalement, une section est consacrée à un bref survol des études économétriques canadiennes utilisant des données d'établissement.

ABSTRACT

This paper reviews the various methods proposed in the literature for estimating economies of scale and identifies the problems associated with each method. The engineering, probabilistic, distribution, survivor and econometric methods are the main approaches proposed in the industrial organization literature and are presented in this paper. In addition, this survey discusses these methods and their limitations in a microeconomic context.

Although all these estimation methods are presented, the emphasis is on the econometric approach. In this context, three types of cost functions are discussed: (i) non-homogeneous and non-homothetic flexible functions; (ii) non-homogeneous and homothetic functions; and (iii) homothetic and homogeneous functions. The consequences and advantages resulting from assuming a homothetic production process are also presented. Finally, a short survey of Canadian econometric studies using plant level data is presented.

TABLE DES MATIÈRES

		Page
1.	Introduction	1
2.	Quelques éléments de théorie	2
3.	Les méthodes d'estimation	5
3.1	La méthode d'ingénierie et les mesures ad hoc	5
3.2	La méthode probabiliste	7
3.3	Les méthodes économétriques	9
3.3.1	Approche primale vs approche duale	9
3.3.2	Fonction de coût non-homogène et non-homothétique	13
3.3.3	Fonction de coût homothétique et non-homogène	15
3.3.4	Fonction de coût homothétique et homogène	22
4.	Bref survol des études microéconomiques	23
4.1	Les études de Fuss-Gupta et de Baldwin-Gorecki	23
5.	Problèmes et critiques	29
5.1	Taux d'utilisation des capacités de production	29
5.2	Régression fallacieuse	30
5.3	Biais de sélection	32
5.4	L'hypothèse de prix des inputs constants	34
5.5	Mesure, diversité et endogénéité de l'output	35
5.6	L'hypothèse d'homothéticité	37
5.7	Petits établissements: salaire et classification	38
5.8	Hétéroscédasticité	38
6.	Conclusion	40

Bibliographie

1. INTRODUCTION

Les économies d'échelle ont une importance majeure lorsqu'il s'agit d'élaborer et d'évaluer des politiques économiques industrielles et commerciales. En effet, les économies d'échelle jouent un rôle clé dans la caractérisation des structures de marché et, de façon similaire, la relation entre le niveau et les coûts de production est une information essentielle pour développer des politiques de concurrence.

D'un point de vue théorique, il est également reconnu depuis longtemps que la prise en compte de certains éléments d'organisation industrielle, tels que les économies d'échelle, a un impact important sur l'analyse du commerce international. Plus récemment, les divers travaux empiriques de Harris et Cox[1], à l'aide d'un modèle d'équilibre général, ont corroboré ce précepte théorique en montrant que l'existence d'économies d'échelle inexploitées avait un impact important sur l'analyse de la formation éventuelle d'une zone de libre-échange entre le Canada et les États-Unis.

Cependant, plusieurs méthodes ont été proposées et utilisées pour estimer les économies d'échelle. L'objectif de ce document est de recenser ces diverses méthodes d'estimation et d'identifier les problèmes empiriques auxquels elles sont confrontées. Les méthodes d'ingénierie, probabiliste, économétrique et les techniques ad hoc basées sur la survie ou la distribution des établissements sont autant d'approches qui ont été proposées dans la littérature d'organisation industrielle. Même si toutes ces méthodes d'estimation sont présentées, nous insistons particulièrement sur l'approche économétrique dans ce texte. De plus, ce survol discute de ces méthodes et de leurs limites à l'intérieur d'un cadre microéconomique.

Le reste de ce document est organisé ainsi: la section 2 présente quelques éléments du cadre théorique. La section 3 procède à un survol des méthodes d'estimation qui ont été proposées dans la

---

1. Voir Harris et Cox (1984), Harris (1984) et Cox et Harris (1985, 1986).

littérature pour évaluer les économies d'échelle. La méthode d'ingénierie et les méthodes dites ad hoc sont d'abord présentées, suivies des méthodes économétriques. Dans ce dernier cas, nous insistons particulièrement sur les conséquences et les avantages résultant de l'hypothèse d'homothéticité du processus de production. La section 4 est consacrée à un court survol des études économétriques et microéconomiques où l'on s'attarde davantage aux études canadiennes récentes. La section 5 tente de faire une synthèse des divers problèmes et critiques associés à l'estimation économétrique des économies d'échelle. En guise de conclusion, nous soulignons les éléments clés du survol et proposons quelques avenues possibles de recherches futures. Finalement, notons que ce survol fait partie d'un projet plus vaste devant déboucher sur des estimations économétriques d'économies d'échelle dans l'industrie manufacturière canadienne.

2. QUELQUES ÉLÉMENTS DE THÉORIE

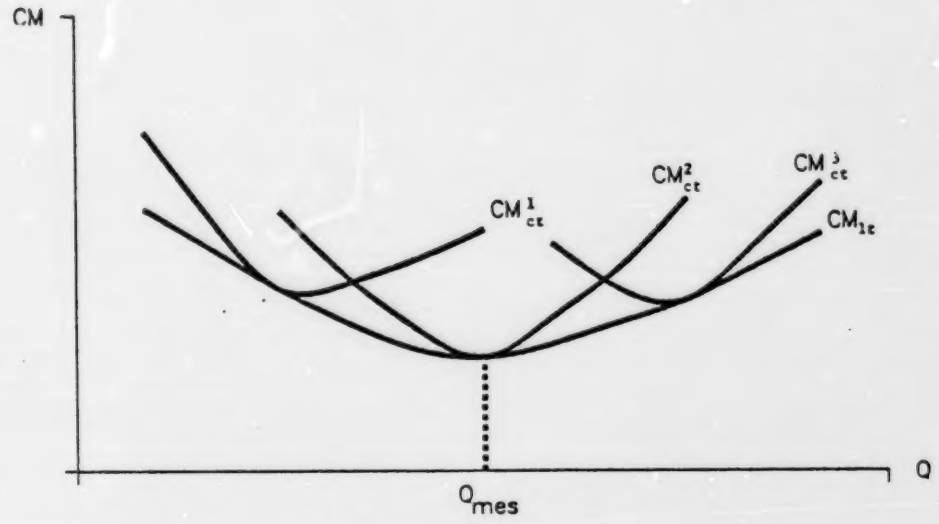
Selon la théorie microéconomique, les économies d'échelle sont déterminées par la forme de la courbe de coût moyen de long terme (CM_{lt}) qui correspond à l'enveloppe des courbes de coût moyen de court terme (CM^{i}_{ct}) comme le montre la figure 1. Chacune de ces courbes de court terme est définie pour un niveau donné du stock d'inputs fixes et quasi-fixes, niveau qui détermine la dimension de la firme.[2] Chaque firme choisit son stock de facteurs fixes et quasi-fixes selon ses anticipations du niveau de production (Q) et les courbes de court et de long termes sont tangentes au point où la production observée et la production anticipée sont égales.

Dans les travaux empiriques, on représente généralement la courbe de coût moyen de long terme à l'aide de deux paramètres de base: (1) le niveau minimum de production optimale (MES) et (2) une approximation ponctuelle de la pente de la courbe. Le niveau minimum de production optimale est défini comme le niveau le plus faible de production où les coûts moyens de long terme sont minimisés et

2. Dans ce document les termes «firme» et «établissement» sont utilisés pour désigner la même entité économique qui correspond, par exemple, à une usine ou une fabrique.

correspond à Q_{mes} sur la figure 1[3]. D'autre part, la pente de la courbe de coût moyen de long terme est approximée par l'augmentation de coût qu'engendre une production sous-optimale donnée (1/2 MES, 1/4 MES etc...) par rapport à une production optimale.

Figure 1: Fonctions de coût de long terme et de court terme



La fonction de coût de long terme est donc la fonction pertinente lorsqu'il s'agit d'évaluer les économies d'échelle. Sur ce point, notons que les coûts totaux de long terme et les coûts variables de long terme sont égaux puisque tous les coûts sont variables à long terme, les facteurs fixes à court terme étant variables à long terme. Le court terme et le long terme sont cependant des concepts relatifs et il est ainsi possible d'imaginer que dans quelques rares secteurs

3. Il s'agit d'une définition théorique. Certaines études empiriques utilisent des définitions légèrement différentes qui sont tout de même basées sur cette définition théorique (voir Fuss et Gupta (1981)).

économiques il existe des coûts fixes, même à long terme[4]. Ensuite, la fonction de coût de long terme est tracée sous l'hypothèse que les prix des inputs[5] et la technologie sont donnés. Tel que souligné par Silberston (1972), l'hypothèse de technologie constante n'implique pas que tous les établissements utilisent la même technique de production. Les établissements de grande dimension peuvent utiliser un processus de production différent de celui utilisé dans les petits établissements, mais ce processus est connu et accessible à tous et il n'est pas exploité par les petits établissements simplement parce qu'il ne correspond pas au processus de production le plus efficient pour ce type d'établissement. De plus, on suppose que l'output de l'industrie, Q, est un produit homogène unique ou un bien de production composite homogène. Ainsi, sous cette hypothèse, les économies d'échelle résultent à la fois d'économies spécifiques à l'établissement et au produit, mais excluent les économies ou déséconomies de diversité.[6]

---

4. Par exemple, dans leur étude sur le secteur de l'agriculture entre 1947 et 1974, Brown et Christensen (1981) supposent que les travailleurs autonomes de fermes et la terre sont respectivement des facteurs quasi-fixes et fixes. De la même façon, l'étude de Caves, Christensen et Swanson (1981) sur les compagnies de transport ferroviaire suppose que les structures et les voies ferrées sont des facteurs quasi-fixes. Dans ce contexte, à long terme, les firmes ne minimisent pas leurs coûts par rapport à l'ensemble des facteurs de production et la fonction de coût total de long terme n'existe pas. Cependant, les firmes minimisent leurs coûts par rapport aux facteurs variables conditionnellement aux niveaux des facteurs fixes et quasi-fixes et il existe une fonction de coût variable de long terme.

5. Ce qui exclut, par conséquent, les économies d'échelle pécuniaires.

6. Nous reviendrons à la section 5 sur l'interaction entre le nombre de produits et la longueur des séries de production d'une part et l'échelle de production d'autre part. On retrouve dans Baldwin et Gorecki (1986a) une discussion de l'interaction de ces «éléments d'échelle» dans le cadre restreint où les économies d'échelle résultent seulement de l'étalement des coûts fixes sur une plus grande échelle de production. Par ailleurs, pour une discussion des sources d'économies d'échelle dans un cadre plus général incluant les facteurs variables, on consultera Silberston (1972), Scherer (1974), Gold (1981) et Green (1985).

Ce cadre theorique etant donne, quel type de donnees doit-on utiliser pour estimer une fonction de coût moyen de long terme: des donnees provenant de coupes transversales ou de series chronologiques? Puisqu'il s'agit d'une fonction microeconomique, l'estimation d'une fonction de coût sur la base de donnees provenant d'une coupe transversale est conceptuellement superieure à celle derivee de donnees en series chronologiques. Comme le soulignent Baldwin et Gorecki (1986, p. 62),

> «The estimates derived from a cross-section are those that pertain to the micro-production function. The time series estimates are those that are relevant to the macro-production function -- the production function written in terms of the sum of the individual establishment variables (outputs, labour, and capital). Sato (1975) has shown that in general the macro-function does not correspond to the micro-function. Thus we would not expect the time series estimates to correspond to micro-estimates. The macro estimates will depend not only on the true micro-production function but also on the distribution of plant sizes and on the changing nature of that distribution over time.»

De plus, les analyses de series chronologiques peuvent aussi refleter des economies d'echelle internes à l'industrie mais externes à la firme. Par consequent, le reste de ce survol discutera des methodes d'estimation et de leurs limites à l'interieur d'un cadre microeconomique.

3. LES MÉTHODES D'ESTIMATION

Même si l'objectif principal de ce texte est de presenter les techniques econometriques utilisees pour evaluer les economies d'echelle, il est utile de presenter rapidement les autres methodes qu'on retrouve dans la litterature d'organisation industrielle.

3.1 La methode d'ingenierie et les mesures ad hoc

La methode d'ingenierie est une approche à laquelle on fait reference dans la plupart des etudes sur les economies d'echelle. Cette methode consiste à evaluer les economies d'echelle à partir d'entrevues

effectuées avec des ingénieurs associés à diverses firmes. Plusieurs considèrent que les études d'ingénierie sont les sources d'information les plus fiables relativement à l'évaluation des économies d'échelle[7]. C'est pourquoi plusieurs auteurs qui utilisent d'autres méthodes qualifient souvent leurs résultats selon la corrélation qu'ils obtiennent avec les estimations d'ingénierie (lorsque ces denières sont disponibles). Cette approche a cependant ses limites. La principale critique qu'on peut adresser à cette méthode est qu'elle ne tient pas compte des coûts de transport. Dans une économie comme celle du Canada où les marchés sont petits et dispersés, les coûts de transport sont importants. Une telle approche essentiellement axée sur la relation intrant-extrant surestimera donc l'impact des économies d'échelle[8]. Il faut aussi prendre en compte la relative rareté de telles estimations: en général, les estimations d'ingénierie ne sont pas récentes et elles ne sont disponibles que pour un nombre limité de secteurs[9].

Une autre mesure du niveau minimum de production optimale qu'on retrouve dans la littérature associée à l'organisation industrielle est dérivée de la distribution de la taille des établissements dans une industrie. Selon cette méthode, le niveau MES est défini comme la taille moyenne du plus petit nombre d'établissements qui compte pour 50 pour cent de la production totale de l'industrie[10].

La technique de survivance est une autre méthode qui a été fréquemment utilisée pour mesurer le niveau MES. Selon cette méthode, les établissements de taille optimale sont ceux qui, entre deux périodes

---

7. Voir Green (1985) et Scherer (1980).

8. On consultera Fuss et Gupta (1981) et Scherer (1980) pour d'autres réserves axées sur la subjectivité des entrevues et l'absence de prise en compte de l'environnement associé au processus de production.

9. Par exemple, les estimations canadiennes les plus récentes sont celles de 1967 présentées dans Scherer (1973) et Scherer et al. (1975) pour quelques industries manufacturières seulement.

10. Cette méthode a été utilisée pour la première fois par Comanor et Wilson (1967).

définies arbitrairement, augmentent la proportion de leurs ventes par rapport à l'ensemble des ventes de l'industrie. On considère alors que ces établissements ont réussi le test crucial du marché: la survie.

On considère souvent, à juste titre, que ces deux dernières techniques de mesure sont purement ad hoc et qu'elles ne peuvent être considérées comme des méthodes de premier plan. En effet, théoriquement, une mesure appropriée du niveau MES doit être basée sur une analyse comparative des coûts moyens entre des établissements de différentes tailles plutôt que sur la distribution de la taille de différents établissements. De plus, d'un point de vue pratique, il est logiquement impossible de penser mesurer le niveau minimum de production optimale à partir de telles méthodes dans une économie comme le Canada où l'on suppose souvent qu'il existe des économies d'échelle inexploitées résultant d'un marché trop petit et d'une concurrence imparfaite. Dans un contexte où l'on suppose que les économies d'échelle ne sont pas exploitées au Canada, l'utilisation de la distribution des établissements américains est probablement plus appropriée que la distribution des établissements canadiens pour évaluer le niveau MES potentiel canadien. Il reste qu'un tel traitement uniforme est insatisfaisant dans un contexte où le degré d'optimalité, en termes d'échelle de production, est variable d'une industrie à l'autre. De plus, ces méthodes ne nous renseignent pas directement sur le désavantage de coûts qu'engendre une production sous-optimale.

3.2 La méthode probabiliste

Plus récemment, Lyons (1980) a proposé une nouvelle méthode pour évaluer le niveau minimum de production optimale. Cette méthode est basée sur une approche probabiliste de la décision de l'entreprise de prendre de l'expansion face à un changement dans la demande de son produit. Cette approche repose sur plusieurs hypothèses relatives à la technologie et au comportement d'une entreprise pouvant être constituée d'un ou plusieurs établissements.

On doit d'abord faire trois hypothèses technologiques afin de spécifier le processus de production de l'entreprise et des établissements qui la composent. Il faut, en premier lieu, supposer que la courbe de

coût unitaire d'un établissement a un forme en L en ce sens qu'elle a une partie horizontale avant de montrer des rendements décroissants. De plus, la partie horizontale de la courbe doit être au moins aussi longue que la distance entre le niveau de production nul et le niveau MES. Cette hypothèse est fondamentale en ce sens qu'elle permet de dériver les hypothèses probabilistes qui suivront relativement au nombre d'établissements que devrait opérer une entreprise étant donné son niveau de production. En dernier lieu, on doit également exclure l'existence d'économies d'échelle reliées à l'opération d'établissements multiples même s'il est possible qu'il y ait des économies internes à l'entreprise.

Ensuite, à partir de cette forme postulée de la courbe de coût moyen, il s'agit de faire des hypothèses probabilistes quant au nombre d'établissements que devrait opérer une entreprise étant donné les niveaux de production et MES. Dans un premier temps, on suppose que la probabilité qu'une entreprise soit constituée d'un seul établissement diminue avec l'écart entre les coûts moyens d'une entreprise avec un établissement (CM_1) et ceux d'une entreprise constituée de deux établissements (CM_2) et que lorsque $CM_1 = CM_2$, la probabilité qu'une entreprise soit constituée d'un seul établissement est égale à 0.5. Deuxièmement, on doit aussi postuler que la probabilité qu'une entreprise soit constituée de trois établissements est nulle avant le niveau de production égal à deux fois MES et augmente ensuite avec le niveau de production.

Sur la base de ces hypothèses, il est possible de dériver le nombre anticipé d'établissements qu'opérera une entreprise. On peut alors démontrer qu'une entreprise aura, en moyenne, 1.5 établissement au niveau de production égal à deux fois le niveau MES. Le niveau MES d'un établissement sera donc égal à la demi de la dimension moyenne des entreprises se retrouvant dans un intervalle de production où le nombre d'établissements moyens par entreprise est de 1.5.

Même si cette méthode est relativement simple et attrayante, elle a cependant le désavantage de reposer largement sur les hypothèses postulées. Comme le montre Lyons (1980), l'estimation du niveau MES que produira une telle méthode surestimera ou sous-estimera le niveau réel

si l'une ou l'autre des hypothèses du modèle n'est pas respectée. Par ailleurs, tout comme dans le cas des mesures dites ad hoc, cette technique ne nous renseigne pas sur la forme de la courbe de coût moyen et l'inefficacité causée par une production sous-optimale.

3.3 Les méthodes économétriques

D'une manière générale, les méthodes économétriques consistent à estimer une fonction de production - l'approche primale - ou une fonction de coût - l'approche duale. L'estimation de l'une ou l'autre de ces fonctions permet d'obtenir tous les paramètres nécessaires à l'évaluation de l'élasticité d'échelle, de la pénalité de coût associée à une production sous-optimale et du niveau minimum de production optimale. Il existe donc deux approches économétriques possibles: l'approche primale et l'approche duale.

3.3.1 Approche primale vs approche duale

Selon l'approche primale du problème de la firme, nous avons la fonction de production

$$[1] \qquad Q = Q(X_1, \ldots\ldots, X_n),$$

où Q est le niveau de production et X_k correspond à la quantité utilisée du facteur de production k. Par définition, l'élasticité d'échelle est donnée par l'expression:

$$[2] \qquad \varepsilon_{QX} = \sum_{i=1}^{n} \frac{\partial Q}{\partial X_i} \cdot \frac{Xi}{Q}$$

Il y a des rendements croissants (décroissants) à l'échelle lorsque ε_{QX} est supérieur (inférieur) à l'unité et le niveau minimum de production optimale est atteint un point où $\varepsilon_{QX} = 1$. Parmi les études microéconomiques qui utilisent cette approche, mentionnons celles de Zellner et Revankar (1969), Griliches et Ringstad (1971), Ringstad (1974), et Baldwin et Gorecki (1986).

Selon l'approche duale, les prix des facteurs de production et le niveau de production sont exogènes et les firmes minimisent leurs coûts de production. Nous avons donc la fonction de coût total:

[3] $$C = C(p_1, \ldots\ldots, p_n, Q),$$

où p_k est le prix du facteur de production k et $C = \sum_i^n X_i p_i$. Par définition, l'élasticité du coût par rapport à la production est donnée par:

[4] $$\varepsilon_{CQ} = \frac{\partial C}{\partial Q} \cdot \frac{Q}{C}$$

et $(\varepsilon_{CQ})^{-1} = \varepsilon_{QX}$.

Il existe cependant plusieurs problèmes d'ordre statistique liés à l'estimation d'une fonction de production ou d'une fonction de coût à l'aide de données en coupes transversales[11]. On peut regrouper ces différents problèmes sous cinq catégories distinctes.

Premièrement, on retrouve généralement une forte collinéarité entre les inputs dans la fonction de production et entre les prix des facteurs dans la fonction de coût, d'où l'existence potentielle d'un problème de multicollinéarité. Selon Baldwin et Gorecki (1986), l'utilisation d'une fonction de coût réduit généralement le problème de multicollinéarité. Il reste que ce type de problème statistique est beaucoup plus important dans les études de séries chronologiques et lorsqu'une forme fonctionnelle flexible est utilisée.

---

11. Même si les problèmes qui suivent sont discutés dans le cadre d'études de coupes transversales, ils se retrouvent aussi à des degrés différents dans les analyses de séries chronologiques. Par ailleurs, pour une analyse comparée des deux approches sous un angle différent, on consultera Baldwin et Gorecki (1986).

Deuxièmement, l'existence des fonctions de production et de coût telles qu'estimées repose respectivement sur des hypothèses technologique et de comportement. D'une part, on suppose que la fonction de production est au moins quasi-concave et, d'autre part, on doit faire l'hypothèse que les firmes minimisent leurs coûts de production. Or, si ces hypothèses ne sont pas respectées, il y a erreur de spécification des fonctions dans les deux cas. Cependant, comme le notent Fuss, McFadden et Mundlak (1978, p. 222) dans un survol des diverses formes fonctionnelles disponibles pour l'analyse du processus de production, «technological and behavioral assumptions which are not widely held to be universal truths, (...) may be widely accepted as plausible for the problem at hand (e.g., «convex technology» or «cost minimizing behavior»)»

Troisièmement, ces méthodes s'exposent à des biais de simultanéité importants. Les quantités des facteurs de production utilisées sont endogènes et déterminées par les prix des facteurs et le niveau de production dans le cas de la fonction de production, tandis que le niveau de production est déterminé par les prix des facteurs et le prix de vente dans le cas de la fonction de coût. Par contre, on peut supposer que les prix des facteurs de production ne sont pas influencés par les producteurs et sont, par conséquent, exogènes. La source de biais de simultanéité est donc multiple dans le cas d'une fonction de production, alors qu'elle est unique lorsqu'on utilise l'approche duale. Ainsi, même s'il est difficile d'évaluer l'impact de chacun de ces biais sur les coefficients estimés, l'approche duale semble être moins affectée par la présence potentielle de biais de simultanéité.

Quatrièmement, un autre type de biais sur les coefficients estimés est sûrement causé par des erreurs dans la mesure des variables explicatives. Ce type de problème affecte aussi bien l'approche primale que l'approche duale et son importance varie d'une banque de données à l'autre. Il est, par conséquent, difficile de traiter de l'impact

potentiel des erreurs de mesure sur l'une ou l'autre des fonctions sans faire référence à une banque de données explicite. Il reste que généralement il ne semble pas qu'une des approches soit beaucoup plus affectée que l'autre par ce type de biais.

Finalement, comme le montrent les études de Ringstad (1974) et Zellner et Revankar (1969), l'utilisation d'une fonction de production non-homogène mais homothétique implique une relation non-linéaire par rapport à Q, la variable dépendante, qui nécessite des méthodes d'estimation non-linéaires coûteuses. À l'inverse, la fonction de coût qui découle d'une fonction de production homothétique peut être estimée directement à l'aide de méthodes linéaires conventionnelles. Dans certains contextes, une telle asymétrie implique un avantage marquée pour l'approche duale. Par exemple, dans le cas canadien, il n'existe pas de données disponibles sur le capital au niveau de l'établissement. Or, comme l'ont démontré Fuss et Gupta (1981), l'hypothèse d'homothéticité du processus de production permet alors d'évaluer l'importance des économies d'échelle sans avoir à utiliser le coût du capital.

En somme, sur la base du troisième et surtout du dernier points, il semble préférable d'opter pour une fonction de coût. Nous allons donc privilégier cette dernière approche dans le reste de cette présentation. Plus précisément, la fonction générale de coût que représente l'équation [3] peut se traduire par diverses formes fonctionnelles spécifiques selon que l'on impose ou non les hypothèses d'homothéticité et d'homogénéité au processus de production. Les trois prochaines sections sont consacrées à la présentation de ces diverses formes.

3.3.2 Fonction de coût non-homogène et non-homothétique

L'approche la plus générale et qui impose le moins de restrictions sur la forme de la fonction consiste à supposer une fonction de coût non-homogène et non-homothétique[12]. Une fonction ayant ces propriétés peut être représentée par diverses formes fonctionnelles[13]. Néanmoins, dans le cadre de ce survol, on peut caractériser ce type de fonction à l'aide d'une forme fonctionnelle translog sans pour autant affecter les conclusions générales de l'analyse. Ce type de fonction a l'avantage de ne pas imposer, a priori, les propriétés d'homothéticité et d'homogénéité et a été utilisé largement ces dernières années pour évaluer le déclin de la croissance de la productivité et la substitution entre les facteurs de production. La forme fonctionnelle translog de l'équation [3] s'écrit:

$$[5] \qquad \ln C = \alpha_o + \alpha_Q \ln Q + \sum_{i=1}^{n} \alpha_i \ln p_i + \tfrac{1}{2} \gamma_{QQ} (\ln Q)^2 + \tfrac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \gamma_{ij} \ln p_i \ln p_j + \ln Q \sum_{i=1}^{n} \gamma_{Qi} \ln p_i$$

La fonction de production sous-jacente à cette fonction de coût sera homothétique si $\gamma_{Qi}=0$ pour tous i et elle sera homogène de degré k si $\gamma_{Qi}=\gamma_{QQ}=0$ et $\alpha_Q=k$. Ainsi, par définition, une fonction de coût homogène est nécessairement homothétique. Ces deux types de fonction ont cependant des propriétés d'échelle très différentes comme nous le verrons dans les sections suivantes.

---

12. Selon Forsund (1975), une fonction de production homothétique à la forme $Q = F[f(X_1,\ldots,X_n)]$ où $f(.)$ est une fonction positive homogène de degré 1, $f(X_1,\ldots,X_n)\in[0, \infty)$ et $F'[f(X_1,\ldots,X_n)] > 0$. D'autre part, une fonction f est homogène de degré r si on peut écrire $f(kX_1,\ldots,kX_n)=k^r f(X_1,\ldots,X_n)$ où k est une constante.

13. On consultera Fuss, McFadden et Mundlack (1978) pour un survol des diverses formes fonctionnelles flexibles.

La première étape consiste à estimer les paramètres de l'équation [5]. Généralement, on obtiendra ces paramètres en estimant simultanément ou en deux étapes, à l'aide de la technique de Zellner, la fonction de coût et les équations des parts de chaque facteur dans le coût total obtenues à partir d'une application du lemme de Shephard[14]. À partir du moment où ces coefficients sont estimés, il est possible, dans une seconde étape, de calculer l'élasticité d'échelle. L'élasticité du coût total par rapport à la production, tel que définie par l'équation [4], est donnée par

$$[6] \qquad \partial \ln C / \partial \ln Q = \alpha_Q + \gamma_{QQ} \ln Q + \sum_{i=1}^{n} \gamma_{Qi} \ln p_i$$

L'inverse de cette expression correspond à l'élasticité d'échelle et le niveau minimum de production optimale est atteint lorsque l'élasticité d'échelle est égale à un.

Si cette forme fonctionnelle a l'avantage d'être flexible théoriquement, elle a cependant le désavantage d'exiger plus de données que certaines autres formes moins générales. De façon plus importante, l'estimation d'une fonction de coût non-homogène et non-homothétique requiert qu'on observe l'ensemble des coûts de production, incluant le coût du capital. Si la fonction de coût est estimée en n'utilisant qu'une partie des coûts comme mesure des coûts totaux, alors les coefficients estimés seront des estimateurs biaisés et non convergents des paramètres recherchés. Or, dans plusieurs pays, dont le Canada, les enquêtes et recensements industriels ne comprennent pas d'information sur les dépenses en capital par établissement[15]. C'est pourquoi, plus

---

14. Dans certains cas, il est possible d'obtenir l'ensemble des paramètres d'échelle en estimant seulement le système d'équations des parts de chaque facteur dans les coûts de production.

15. Par exemple, au Canada le recensement des manufactures fournit des statistiques sur la valeur des livraisons, les inventaires, le travail, les matières intermédiaires, l'énergie et d'autres caractéristiques par établissement. Par contre, le recensement ne contient pas d'information sur le capital.

souvent qu'autrement, les études qui utilisent cette forme fonctionnelle sont effectuées pour des secteurs particuliers ou à l'aide de données en séries chronologiques. Il est donc important d'examiner l'impact qu'ont les hypothèses d'homothéticité et d'homogénéité du processus production sur l'estimation des économies d'échelle.

3.3.3 Fonction de coût homothétique et non-homogène

Une alternative à la fonction de production non-homothétique et non-homogène est l'utilisation d'une fonction de production homothétique, mais non-homogène. La principale caractéristique de ce type de fonction est de supposer que les proportions des facteurs de production utilisées sont indépendantes du niveau de production. La fonction de coût total qui est dérivée d'une telle fonction de production a donc la propriété de séparabilité entre les prix des facteurs de production et la quantité produite. De plus, le travail novateur de Fuss et Gupta (1981) a démontré que l'existence d'un processus de production homothétique permettait d'identifier les paramètres d'échelle de la fonction de coût total via l'estimation de la partie variable de la fonction de coût total de long terme.

Par définition, la fonction de coût qui découle d'une fonction de production homothétique prend la forme:

$$[7] \qquad C(p,Q) = h(p)T(Q)$$

où p est un vecteur qui contient les prix des inputs, Q est la quantité produite, C(.) est la fonction de coût total et T(Q) est la fonction du sentier d'expansion de la production.

On peut démontrer la propriété d'indépendance entre les proportions des facteurs de production et le niveau de production directement à partir de l'équation [7]. En utilisant le lemme de Shephard, les fonctions de demande des inputs i et j sont:

$$[8] \qquad v_i = \partial C/\partial p_i = (\partial h(p)/\partial p_i)T(Q) = h_i(p)T(Q)$$

$$[9] \qquad v_j = \partial C/\partial p_j = (\partial h(p)/\partial p_j)T(Q) = h_j(p)T(Q)$$

ainsi,

$$[10] \qquad v_i/v_j = h_i(p)/h_j(p)$$

L'équation [10] correspond à la propriété fondamentale d'une fonction de coût homothétique: les proportions des facteurs de production utilisés sont indépendantes du niveau de production.

Si le coût du capital est maintenant exclu des coûts totaux, on obtient la fonction de coût «variable» de long terme[16]:

$$[11] \qquad CV(p,Q) = \sum_{i \neq k} p_i v_i = \sum_{i \neq k} p_i h_i(p) T(Q) = g(p)T(Q)$$

où $g(p) = \sum_{i \neq k} p_i h_i(p)$ et k représente le capital. À prix constants, cette fonction correspond à une approximation parfaite du coût total puisque $CV(p,Q)/C(p,Q) = g(p)/h(p)$; c'est-à-dire que le rapport entre les coûts «variables» et les coûts totaux est indépendant du niveau de l'output. Par conséquent, comme le notent Fuss et Gupta (1981), les équations [7] et [11] contiennent le même sentier d'expansion de la production, T(Q), et l'une ou l'autre peut être utilisée pour évaluer les économies d'échelle. À l'inverse, on ne peut retrouver un tel résultat dans le

16. En théorie microéconomique, les coûts variables sont généralement définis par opposition aux coûts fixes dans un contexte de court terme. Ce n'est pas dans ce sens que nous utilisons cette expression ici, puisqu'on se situe dans une perspective de long terme. Dans le présent contexte, les coûts «variables» correspondent plutôt aux coûts totaux moins les coûts provenant de l'utilisation des services du capital. Alors que la fonction de coût variable de court terme met en relation les coûts variables avec les prix des inputs variables, les quantités des inputs fixes et le niveau de production, la fonction de coût «variable» de long terme, quant à elle, met en relation les coûts «variables» avec l'ensemble des prix des inputs et le niveau de production.

cas d'une fonction non-homothétique où il y a interaction entre le prix des inputs et le niveau de production dans la fonction h de l'équation [7] [17].

En somme, les travaux de Fuss et Gupta (1981) montrent que l'hypothèse d'homothéticité du processus de production est une condition nécessaire et suffisante pour obtenir une description complète des économies d'échelle associées à l'ensemble du processus de production à partir d'une fonction de coût «variable» de long terme.

En supposant que la forme fonctionnelle spécifique de la fonction g(p) est connue, il ne reste qu'à spécifier la forme fonctionnelle de T(Q) pour que l'équation [11] soit estimable. Une sélection des divers sentiers d'expansion proposés dans la littérature et les fonctions de coût correspondantes sont présentés au tableau 1. Les fonctions implicites de production sous-jacentes aux fonctions de coût sont aussi présentées pour trois des cinq formes fonctionnelles. Comme nous l'avons souligné précédemment, l'étude du tableau 1 montre que l'estimation de ces fonctions de production ne peut être réalisée à l'aide de méthodes linéaires usuelles puisque certains paramètres à estimer (γ_1, γ_2) sont associés à la variable dépendante Q. Par conséquent, l'estimation directe de ces fonctions implicites de production requiert des méthodes non-linéaires coûteuses, ce qui représente un désavantage par rapport à l'approche duale.

---

17. Une fonction de coût total générale non-homothétique peut s'écrire $C(p,Q)=h(p,Q)T(Q)$.

Tableau 1: Quelques fonctions homothétiques proposées dans la littérature

Étude	$T(Q)$[1]	Fonction de coût[2]	Fonction de production
DE BRANBANDER - VANLOMMEL (1978) GUPTA (1968)	$a + bQ + cQ^2$	$(C/Q) = a/Q + b + CQ$	-
GUPTA-FUSS (1979) FUSS-GUPTA (1981)	$Qe^{a + bQ + c/Q}$	$\ln(C/Q) = a + bQ + c/Q$	-
NERLOVE (1963)	$\beta_0(\ln Q) + \beta_1(\ln Q)^2$	$\ln C = \beta_o \ln Q + \beta_1(\ln Q)^2 + \sum_i \alpha_i \ln p_i$	$\ln Q + \gamma_1(\ln Q)^2 = \sum_i \gamma_i \ln X_i$
RINGSTAD (1974)	$\beta_0(\ln Q) + \beta_1(\ln Q)^2 + \beta_2 Q$	$\ln C = \beta_o \ln Q + \beta_1(\ln Q)^2 + \beta_2 Q + \sum_i \alpha_i \ln p_i$	$\ln Q + \gamma_1(\ln Q)^2 + \gamma_2 Q = \sum_i \gamma_i \ln X_i$
ZELLNER-REVANKAR (1969)	$\beta_0(\ln Q) + \beta_1 Q$	$\ln C = \beta_o \ln Q + \beta_1 Q + \sum_i \alpha_i \ln p_i$	$\ln Q + \gamma_1 Q = \sum_i \gamma_i \ln X_i$

1. Pour les études de Nerlove (1963), Ringstad (1974) et Zellner-Revankar (1969), il s'agit de ln (T(Q)).

2. Les coefficients β_i peuvent s'exprimer en fonction des coefficients γ_i et vice versa. Pour une dérivation explicite de ces relations dans le cas des fonctions de Nerlove, Ringstad et Zellner-Revankar, on consultera Ringstad (1974).

Comme le notent Fuss et Gupta (1981), il y a cependant peu de chances que l'hypothèse d'homothéticité du processus de production soit vérifiée dans les faits. En effet, on suppose généralement que les firmes de plus grande dimension ont une production plus intensive en capital. La question est de savoir jusqu'à quel point l'hypothèse imposée ne respecte pas la réalité. Comme nous l'avons noté plus haut, sur la base d'une fonction de coût translog, la restriction d'homothéticité correspond à imposer une valeur nulle aux coefficients $\gamma_{Qi}(i=1,....,n)$ dans l'équation [5]. Toujours sur cette base, le biais causé par l'imposition de l'hypothèse d'homothéticité sur l'élasticité d'échelle correspond donc au terme $\Sigma\gamma_{Qi}\ln p_i$ selon l'équation [6].

Théoriquement, tel que le soulignent Fuss et Gupta, la condition d'agrégation d'Engel, selon laquelle $\Sigma p_i v_i/C = 1$, implique $\Sigma\gamma_{iQ} = 0$ même lorsque le processus de production n'est pas homothétique. Ainsi, d'un point de vue théorique, il semble raisonnable de penser que le biais d'homothéticité est faible.

Par ailleurs, d'après Fuss et Gupta, la seule information empirique disponible sur l'ampleur de ce biais, dans le cas canadien, se retrouve dans l'étude de Denny, Fuss et Waverman (1979). Ces derniers estiment une fonction de coût translog générale pour les vingt industries manufacturières canadiennes à deux chiffres sur la période 1962-1973. Leurs résultats suggèrent que le biais varie entre 0.001 et 0.02 par rapport à une élasticité-coût ($\partial\ln C/\partial\ln Q$) variant entre 0.8 et 1.2 selon l'industrie. Précisons ici que ces résultats indiquent le biais d'homothéticité que causerait l'exclusion des termes croisés entre le niveau de production et les prix si les coefficients des autres variables n'étaient pas affectés par des biais de spécification résultant de l'exclusion de ces termes croisés. Évidemment, s'il existe une corrélation entre les termes croisés et les autres variables (ce qui est à toutes fins utiles assurée), alors l'omission des $\ln Q\ln p_i$ $(i=1,...,n)$ créera un biais de spécification sur les coefficients des autres variables indépendantes. Il reste que ce biais de spécification peut aussi bien augmenter, réduire ou ne pas affecter

le biais d'homothéticité affectant l'estimation des économies d'échelle. Néanmoins, ces résultats tirés de l'étude de Denny, Fuss et Waverman (1979) incitent Fuss et Gupta (1981, p.128) à conclure que:

> «the homotheticity restriction, which permits identification of scale effects, has a high benefit/cost ratio; in the sense that identification is obtained at the cost of a small estimation bias, at least in the case of Canadian manufacturing.»

Cependant, deux réserves importantes doivent être soulignées quant à cette conclusion sur l'impact de l'hypothèse d'homothéticité sur les estimations d'économies d'échelle. D'abord, le fait de supposer un processus de production homothétique alors qu'il ne l'est pas peut biaiser indirectement les économies d'échelle estimées si cette hypothèse est utilisée pour estimer une fonction de coût «variable» plutôt qu'une fonction de coût total. Si les travaux de Fuss et Gupta nous renseignent sur l'impact potentiel de l'exclusion des termes du second ordre sur l'estimation des économies d'échelle lorsque le processus de production n'est pas homothétique, il ne nous permet pas d'évaluer l'impact d'exclure le coût du capital des coûts totaux. Or, il est possible que l'exclusion du capital ait un impact beaucoup plus important sur l'estimation des économies d'échelle que la simple exclusion des termes du second ordre si le processus de production n'est pas homothétique. Ensuite, tel que nous l'avons noté plus haut, les résultats tirés d'études en séries chronologiques sont discutables à cause des divers biais d'agrégation possibles. Par conséquent, il n'y a pas de raison de croire que les résultats macroéconomiques tirés de l'étude de Denny, Fuss et Waverman s'appliquent au niveau des fonctions microéconomiques correspondantes (Sato (1975)).

Pour examiner l'évidence empirique existant au niveau microéconomique relativement à l'exclusion du capital, il est utile d'illustrer de façon différente ce qu'implique l'hypothèse d'homothéticité pour le processus de production. En suivant la même procédure d'exclusion que celle utilisée pour obtenir l'équation [11], on peut écrire:

$$[12] \qquad C_i(p,Q) = g_i(p)T(Q)$$

où C_i est le coût associé au(x) facteur(s) de production i. Donc, si le processus de production est homothétique, l'observation des coûts associés à un seul facteur de production est suffisant pour évaluer les économies d'échelle spécifiques à l'ensemble du processus de production. En d'autres termes, cela implique que l'élasticité de la partie des coûts spécifiques à chaque input par rapport à l'output est identique pour tous les inputs:

$$[13] \qquad \varepsilon_{CQ} = \frac{\partial \ln C_i}{\partial \ln Q} = \frac{\partial \ln C_j}{\partial \ln Q} = \frac{\partial \ln C}{\partial \ln Q}$$

Or, un tel résultat est hautement improbable comme l'indiquent les résultats de Haldi et Withcomb (1967). Ceux-ci arrivent à la conclusion qu'il existe des économies d'échelle importantes dans le cas du capital et du travail, faibles dans le cas de l'énergie et nulle dans le cas des matières premières. Ainsi, par exemple, l'exclusion du coût des matières premières des coûts totaux impliquera un biais positif important sur les économies d'échelle. Par contre, dans le cas d'une fonction de coût «variable» de long terme, il s'agit de savoir jusqu'à quel point l'élasticité d'échelle associée au coût du capital seulement diffère ou s'approche de l'élasticité d'échelle associée aux coûts «variables».

Pour ce faire, on peut se référer aux résultats empiriques obtenus par Haldi et Withcomb (1967). Ils ont estimé la fonction,

$$[14] \qquad C_i = A_i Q^{b_i}$$

pour les facteurs variables et le capital séparément à l'aide de données tirées d'études d'ingénierie. Les facteurs variables correspondent aux matières premières et au travail (incluant la supervision, la gestion et l'entretien). En prenant la moyenne des estimations présentées par intervalle, b_i est égal à .68 dans le cas des coûts variables et à .64 pour les coûts associés au capital. Ainsi, sur la base de cette étude,

l'exclusion des coûts du capital, basée sur l'hypothèse d'homothéticité du processus de production, ne semble pas causer un biais négatif important sur l'estimation des économies d'échelle.

Évidemment, il est difficile de se baser sur une seule étude, datant de plusieurs années par surcroît, pour tirer des conclusions définitives. Malheureusement, à notre connaissance il n'existe pas d'autres études empiriques et microéconomiques qui puissent nous informer sur l'importance du biais d'homothéticité.

3.3.4 Fonction de coût homothétique et homogène

Une dernière forme fonctionnelle possible consiste à supposer une fonction de production ou de coût homothétique et homogène. Pour fins d'illustration, on peut représenter cette famille de forme fonctionnelle à l'aide d'une fonction de type Cobb-Douglas. Alors qu'une fonction de coût translog correspond à une approximation du second ordre d'une fonction de coût générale, la fonction de coût Cobb-Douglas constitue une approximation du premier ordre de la même fonction. Une fonction de coût total Cobb-Douglas s'écrit:

$$[14] \qquad \ln C = \alpha_o + \alpha_Q \ln Q + \sum_{i=1}^{n} \alpha_i \ln p_i$$

où α_Q correspond à l'élasticité du coût par rapport à la production. La fonction de production sous-jacente à cette fonction de coût possède les propriétés d'homothéticité et d'homogénéité de degré α_Q.

Ce type de fonction de production a été utilisé par Baldwin et Gorecki (1986) dans le but d'évaluer l'impact des économies d'échelle sur l'écart de productivité entre les industries manufacturières canadienne et américaine à un niveau de désagrégation à quatre chiffres pour les années 1970 et 1979. Parmi les nombreuses autres études ayant utilisé cette forme fonctionnelle, mentionnons les études de Griliches (1967) et de Griliches et Ringstad (1971).

Ce type de forme fonctionnelle a cependant le désavantage important d'empêcher l'élasticité d'échelle de varier avec le niveau de production. Ainsi, par définition, une fonction de coût unitaire homothétique et homogène n'a pas de point minimum ce qui exclut, par exemple, l'estimation de courbe de coût moyen en U. Cette caractéristique de la fonction Cobb-Douglas la rend donc inadmissible ou inutile pour une évaluation du niveau MES. De la même façon, l'écart de coût en pourcentage qu'indique une telle fonction entre deux niveaux de production est trompeur puisqu'il correspond en fait à une moyenne pour l'ensemble du spectre observé de l'output.

4. BREF SURVOL DES ÉTUDES MICROÉCONOMIQUES

Comme dans la plupart des domaines de la recherche économique, l'abondance de données agrégées relativement aux données individuelles a fait en sorte que la majorité des efforts empiriques pour estimer les économies d'échelle a été faite à partir de données en séries chronologiques. Au Canada, les études de Zohar (1982), Rao et Preston (1983) et Daly et Rao (1986) sont des exemples de ce type d'étude dans l'industrie manufacturière.

De leur côté, les études microéconomiques sont surtout effectuées dans des secteurs réglementés puisque ces secteurs respectent vraisemblablement l'hypothèse d'exogénéité du niveau de production sous-jacente à la théorie duale de la production. On retrouve dans Jorgenson (1984) un survol de plusieurs de ces études. Il existe aussi un certain nombre d'études microéconomiques qui tentent d'évaluer les économies d'échelle dans les industries manufacturières à un niveau d'agrégation relativement fin. Parmi ces études, mentionnons celles de Gupta (1968) - Inde -, Griliches et Ringstad (1971) - Norvège -, De Brabander et Vanlommel (1978) - Belgique - et Corbo et Meller (1979) - Chili. Ces derniers utilisent une fonction de production translog et sont les seuls à utiliser une forme fonctionnelle flexible. Griliches et Ringstad (1971) utilisent une fonction de production Cobb-Douglas, alors que les autres études utilisent des fonctions de coût homothétiques. Pour ce qui est du Canada, les études utilisant des données microéconomiques sont plutôt rares: les études de Fuss et Gupta (1981) et Baldwin et Gorecki (1986) sont seules dans cette catégorie.

4.1 Les études de Fuss et Gupta et de Baldwin et Gorecki

Fuss et Gupta (1981) estiment une fonction de coût unitaire homothétique pour 91 industries manufacturières de la classification industrielle à 3 chiffres. Ils n'utilisent pas directement des données internes aux établissements, mais plutôt des moyennes de ces données regroupées selon la taille des établissements. Ces données groupées étant rarement disponibles pour plus de huit catégories, ils ont réuni les données des années 1965, 1968, 1969 et 1970 pour augmenter le nombre de degrés de liberté associé à chaque régression. C'est l'approche duale qui est utilisée et la fonction de coût moyen estimée est la suivante:

$$[15] \qquad \ln(C/Q) = \alpha_1 + \alpha_2 Q + \alpha_3 Q^{-1} + \sum_{t=2}^{4} B_t D(t) + u$$

où C est le coût de production excluant le coût du capital, Q est le niveau de production mesuré par la production brute, $\alpha_i (i = 1, 2, 3)$ sont des coefficients à estimer, D_t comprend les variables dichotomiques spécifiques aux temps, B_t est le vecteur des coefficients associés à ces variables dichotomiques et u est un terme d'erreur aléatoire centré sur zéro. Puisque Fuss et Gupta font l'hypothèse que le prix de chaque facteur de production ne varie pas d'un établissement à l'autre dans une même industrie, ces prix sont exclus de la partie déterministe de la fonction de coût. De plus, pour corriger l'hétéroscédasticité causée par l'inégalité du nombre d'établissements dans chaque agrégat, la méthode des moindres carrés pondérés est utilisée plutôt que celle des moindres carrés ordinaires.

Trois régressions sont effectuées pour chaque industrie sur la base de l'équation [15]: une première incluant les variables Q et Q^{-1} (i.e. l'équation [15]); une seconde excluant seulement Q; et une dernière excluant seulement Q^{-1}. L'une de ces régressions est considérée comme représentant la vraie courbe de coût unitaire d'une industrie sur la

base du critère de minimisation de l'écart-type de la régression.[18] Cette procédure utilisée par Fuss et Gupta est asymétrique en ce sens qu'elle permet à la fonction de coût d'exhiber des rendements à l'échelle croissants ou décroissants mais l'empêche de produire des rendements constants. Une procédure symétrique peut être obtenue simplement en contraignant les paramètres α_2 et α_3 à zéro. On obtient alors un quatrième modèle qui peut ainsi être confronté aux trois premiers précédemment exposés.

De leur côté, Baldwin et Gorecki (1986) estiment plutôt une fonction de production dans environ 160 industries manufacturières de la classification industrielle à 4 chiffres pour les années 1970 et 1979. Alors que Fuss et Gupta utilisent des données groupées, Baldwin et Gorecki utilisent quant à eux des données internes aux établissements provenant du recensement des manufactures. Ils estiment directement une fonction de production Cobb-Douglas qui, par définition, prend la forme:

$$[16] \qquad \ln VA = \alpha_1 + \alpha_2 \ln L + \alpha_3 \ln K + u$$

où VA est le niveau de production mesuré par la valeur ajoutée, L est la quantité de travail mesurée par le nombre total d'employés et K est une approximation des services du capital basée sur les dépenses en énergie.

Les résultats obtenus par Fuss et Gupta (1981) pour la fin des années soixante suggèrent qu'il existe des rendements à l'échelle croissants dans les 91 industries manufacturières analysées. Pour sa part, l'étude de Baldwin et Gorecki (1986) indique qu'il existe des rendements croissants, constants et décroissants dans respectivement 65%, 33% et 2% des industries manufacturières, ou encore pour 78%, 21% et 1% de la valeur ajoutée de l'ensemble du secteur manufacturier en 1970. Leurs résultats pour 1979 indiquent par ailleurs des rendements croissants, constants et décroissants dans 52%, 46% et 2% des industries manufacturières, ou 64%, 32% et 4% de la valeur ajoutée totale du

---

18 Il est ainsi possible que les coefficients associés à la production ou à son inverse ne soient pas significativement différents de zéro aux niveaux habituels de confiance dans la régression finalement choisie pour représenter la fonction de coût de l'industrie.

secteur manufacturier.[19] L'étude de Baldwin et Gorecki laisse donc voir une réduction sensible de l'importance des économies d'échelle dans le secteur manufacturier au cours des années 70.

Même si, comme nous l'avons noté plus haut, la procédure utilisée par Fuss et Gupta exclut par définition l'existence d'industrie à rendements constants, les résultats obtenus par ces derniers semblent indiquer des économies d'échelle plus importantes que l'étude de Baldwin et Gorecki. Une telle conclusion est cependant trompeuse puisque si les résultats présentés plus haut donnent une bonne idée de l'existence ou de l'absence d'économies d'échelle d'une industrie à l'autre, ils n'indiquent pas l'ampleur de ces économies, telle que mesurée par la pente de la courbe de coût. Une investigation de l'importance des économies d'échelle peut se faire en calculant les moyennes pondérées des élasticités d'échelle dérivées de ces études. Sur cette base, en utilisant le niveau de production comme poids, la conclusion précédente est inversée: le niveau des économies d'échelle est beaucoup plus considérable selon l'étude de Baldwin et Gorecki. La moyenne pondérée des économies d'échelle est de l'ordre de 1.22 dans le cas de l'étude de Baldwin et Gorecki comparativement à 1.06 selon les résultats de Fuss et Gupta[20].

Comment réconcilier ces résultats divergents? A priori, il existe au moins quatre différences entre ces deux études qui peuvent potentiellement expliquer cette divergence. En effet, les deux études diffèrent au niveau: (1) de la forme fonctionnelle; (2) de l'année d'estimation; (3) du type de données utilisées; et (4) de la mesure du niveau de production.

---

19. Ces résultats sont basés sur des tests statistiques t standard où le seuil de confiance est fixé à .10.

20. Il s'agit d'une moyenne pondérée pour les 91 industries de l'échantillon utilisée par Fuss et Gupta. Les élasticités d'échelle par industrie se retrouvent au tableau D-1 de Baldwin et Gorecki, p.283. Nous utilisons la valeur des ventes et la valeur ajoutée en 1970 comme poids pour Fuss et Gupta et Baldwin et Gorecki respectivement.

Si l'utilisation de données individuelles plutôt que des données groupées devrait améliorer la précision des estimations, il serait surprenant que cela ait un effet aussi substantiel sur l'estimation des économies d'échelle. En fait, il est difficile d'imaginer que l'une ou l'autre des trois premières différences puissent avoir un impact aussi important sur l'estimation des économies d'échelle. Par contre, d'un point de vue théorique, l'utilisation par Baldwin et Gorecki de la valeur ajoutée comme mesure de l'output plutôt que la production brute peut expliquer le sens et l'ampleur de l'écart entre les estimations des économies d'échelle obtenues par ces deux études. En effet, l'exclusion des matériaux de l'output et des coûts devrait élever de façon substantielle les estimations d'élasticité d'échelle puisque, a priori, il n'y a pas de raison de croire qu'il puisse exister des économies d'échelle technique au niveau des matériaux. De plus, les matériaux constituent généralement une part importante des coûts totaux.

On peut vérifier cette hypothèse - valeur ajoutée/production brute - à l'aide des résultats obtenus par Baldwin et Gorecki (1986) avec la régression supplémentaire suivante:

$$[17] \qquad \ln Q = \beta_1 + \beta_2 \ln L + \beta_3 \ln K + \beta_4 \ln M + u$$

où M correspond aux coûts des matériaux, Q est la production brute et les autres variables ont été définies précédemment. L'équation [17] a été estimée pour un sous-échantillon de 8 industries et les élasticités d'échelle résultant de l'estimation des équations [15] à [17] pour ces industries sont consignées au Tableau 2.

Comme le montre le Tableau 2, les régressions effectuées par Baldwin et Gorecki (1986) avec une fonction de production brute produisent des élasticités d'échelle qui sont toujours substantiellement plus faibles que celles dérivées d'une fonction de valeur ajoutée. Les résultats obtenus avec une fonction de production brute, par Baldwin et Gorecki, se rapprochent en effet beaucoup plus des estimations de Fuss et Gupta (1981) que de celles résultant d'une fonction de valeur ajoutée. Cet exercice confirme donc l'importante différence existant entre les économies d'échelle liées aux facteurs de production primaires et secondaires.

Tableau 2: Élasticités d'échelle estimées

Code[1]	Industrie	Baldwin-Gorecki 1970 VA (Eq. 16)	Baldwin-Gorecki 1970 Q (Eq. 17)	Fuss-Gupta 1965-1970 Q (Eq.15)
1072	Boulangerie	1.14	1.07	1.10
1081	Fabricants de confiserie	1.16	1.09	-
1832	Tissage, moulinage et filage	1.06	0.94	1.01
2513	Scieries et at. de rabotage	1.43	1.01	1.16
2860	Imprimerie commerciale	1.15	1.09	1.05
3042	Emb. et mat. des métaux	1.22	1.07	1.04
3320	Gros appareils	1.15	0.98	1.03
3360	Équipement électrique ind.	1.04	1.03	1.00
	Moyenne (excluant 1081)	1.17	1.03	1.06

1. Il s'agit des codes de la classification industrielle de 1970. Les codes de la classification industrielle de 1960 utilisés par Fuss et Gupta correspondant à ces codes de 1970 sont dans l'ordre; 129, 131, 201, 2513, 286, 304, 332 et 336.

Source: Baldwin et Gorecki (1986), tableau C-7, p. 327 et tableau D-1, p. 334.

Cela étant dit, laquelle de ces deux approches est appropriée dans le cadre d'une analyse sectorielle? Notons d'abord que l'utilisation d'une fonction de valeur ajoutée requiert que les inputs intermédiaires et la valeur ajoutée soient séparables. Selon Denny et May (1977) et Delorme et Lester (1986), cette hypothèse n'est pas vérifiée empiriquement dans les industries manufacturières canadiennes. L'estimation d'une fonction de production ou de coût basée sur un concept de production brute apparaît donc comme une approche plus robuste. De plus, du point de vue de la firme, les coûts pertinents sont les coûts totaux et les coûts associés aux facteurs de production primaires ne représentent qu'une partie de ces coûts. S'il est valable

d'utiliser la valeur ajoutée à un niveau fortement agrégé de l'économie pour éviter le double comptage, cela n'apparaît pas approprié à un niveau industriel relativement désagrégé.[21]

5. PROBLÈMES ET CRITIQUES

Les analyses dont l'objectif est d'estimer une fonction de coût et les économies d'échelle qui en résultent sont confrontées à de nombreux problèmes théoriques et empiriques. Dans cette section, nous tenterons quant à nous de résumer les principaux problèmes empiriques liés aux estimations économétriques d'économies d'échelle.[22] Certains éléments de solution ou de correction sont aussi présentés.

5.1 Taux d'utilisation des capacités de production

Comme nous l'avons noté plus haut, les économies d'échelle sont déterminées par la forme de la courbe de coût moyen de long terme qui correspond à l'enveloppe des courbes de coût moyen de court terme. Or, on n'estimera réellement la fonction de long terme que si les établissements observés produisent à leurs niveaux de production planifiés ou anticipés: c'est-à-dire, s'ils produisent aux points où leurs courbes de court terme sont tangentes à la courbe de long terme sur la figure 1. Si une telle situation correspond au meilleur des mondes pour un économètre, elle ne correspond probablement pas à la réalité. Il est en effet facile d'imaginer que certains établissements produisent à des niveaux différents des niveaux anticipés. Ainsi, une

---

21. Évidemment, une plus grande exploitation des économies d'échelle (associées aux inputs primaires) aura, par ricochet, l'effet de réduire les coûts des matériaux, mais, en soi, cette réduction des coûts ne sera d'aucune façon reliée à l'échelle de production: la baisse du coût des matériaux ne sera pas moins importante dans les petits établissements.

22. Pour un survol original, inter alia, des problèmes associés à l'analyse théorique des économies d'échelle, on consultera Gold (1981).

firme se situera sur sa courbe de court terme (CM^{i}_{ct}), mais pas nécessairement au point de tangence entre sa courbe de court terme et la courbe de long terme de l'industrie (CM_{lt}) sur la figure 1. Dans ce contexte, la courbe estimée se situera à l'intérieur de la courbe réelle de coût moyen.

Quelques modèles intégrant le taux d'utilisation des capacités de production ont déjà été proposés dans la littérature.[23] Cependant, ces modèles s'appliquent essentiellement aux données en séries chronologiques puisqu'ils nécessitent la mesure d'un taux d'utilisation «normal» des capacités de production. Une transposition de ce type de modèle aux données d'établissements nécessiterait l'observation de plusieurs années pour un même établissement, situation peu courante.

Une autre méthode, proposée par Hammond (1986), consiste à contraindre les termes d'erreurs à être uniquement positifs. En effet, selon Hammond, en partant de l'hypothèse que certains établissements sont en équilibre de long terme alors que d'autres sont plutôt en équilibre de court terme, c'est la frontière inférieure des observations de coût qui permet d'identifier la fonction de coût de long terme.

Finalement, dans une analyse de coupe transversale, le choix d'une année où l'économie ne subit pas de chocs importants constitue sûrement un prérequis simple et minimal.

5.2 Régression fallacieuse

L'estimation d'une fonction de coût moyen de long terme peut aussi souffrir d'un problème de «regression fallacy» mis en évidence par Friedman (1955).[24] L'argument de Friedman est basé sur le fait qu'on ne peut observer le coût de production effectif lorsque le niveau de production observé est différent du niveau planifié. En accord avec

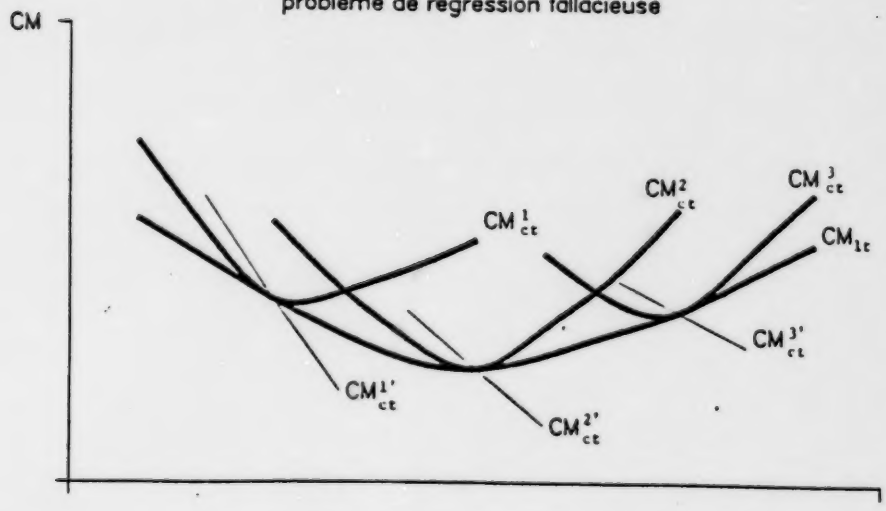
---

23. Voir Fuss et Waverman (1985), par exemple.

24. On retrouve plusieurs interprétations du problème de régression fallacieuse dans la littérature. La présentation qui suit est basée sur Friedman (1955) et l'analyse qu'en fait Borts (1960).

la théorie, les firmes minimisent leurs coûts variables à court terme, mais ce processus de minimisation n'est pas reflété par les données sur les coûts généralement utilisées et disponibles. Par exemple, des changements dans la partie variable des coûts associés au capital- (réparation, entretien ou même remplacement) et le nombre d'heures travaillées ne sont pas toujours reflétés dans les statistiques disponibles. Par conséquent, les firmes avec un niveau de production supérieur (inférieur) au niveau planifié auront des coûts moyens observés inférieurs (supérieurs) aux coûts réel de long terme et de court terme. En d'autres termes, même si les firmes se situent effectivement sur les courbes de court terme (CM^{i}_{ct}), on ne les observent pas sur celles-ci, mais plutôt sur les courbes hybrides ($CM^{i'}_{ct}$) sur la figure 2. Quel est l'effet final de ce problème sur la pente de la courbe de coût moyen? Sous l'hypothèse d'indépendance entre les distributions des niveaux de production planifié et transitoire, les firmes avec un niveau de production supérieur à la production moyenne de l'ensemble des firmes de l'industrie auront, en moyenne, une composante transitoire positive et inversement pour les autres firmes dont le niveau de production est inférieur à la production moyenne.[25] Ainsi, la pente de la courbe de coût moyen sera biaisée négativement.

Figure 2: Illustration du problème de régression fallacieuse



25. Pour une démonstration, on consultera Johnston (1958).

Whalley (1984) et Hazeldine (1984) affirment que les résultats de Fuss et Gupta (1981), et, par ricochet, les études utilisant une méthode et des données semblables, peuvent être le reflet d'un tel problème et non d'économies d'échelle réelles. Cette critique mérite, à tout le moins, d'être nuancée. En effet, comme le notent Baldwin et Gorecki (1986), l'utilisation de données groupées par Fuss et Gupta (1981) élimine, du moins en partie, le biais causé par ce problème selon les résultats empiriques obtenus par Borts (1960). De plus, en supposant que la majeure partie de l'écart entre les coûts réels et les coûts observés provient de la partie variable des coûts en capital, l'estimation d'une fonction de coût «variable» ne sera pas affectée significativement par un tel problème.

5.3 Biais de sélection

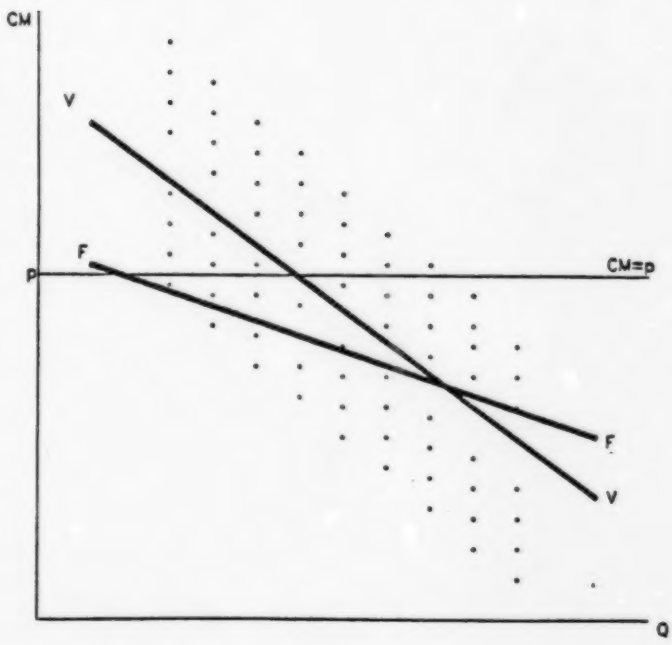
Une des caractéristiques fondamentales des analyses économétriques est qu'elles sont contraintes par les données coût/output observées. Par conséquent, les estimations économétriques d'économie d'échelle sont potentiellement affectées par des biais de sélection puisque la distribution observée de l'output correspond probablement à une distribution doublement tronquée de la vraie distribution dans la plupart des industries.

Au niveau de la partie supérieure de la distribution, il est théoriquement difficile d'observer des établissements dans la partie ascendante de la courbe de coût moyen. De plus, dans plusieurs industries, les plus grands établissements se situent bien en deçà du niveau optimal de production. Ainsi, la troncature de la partie supérieure de la distribution de l'output implique une sous-estimation des économies d'échelle et surtout du niveau minimum de production optimale.

Plus récemment, Miller (1977) a aussi proposé l'existence d'un biais de sélection causé par la troncature de la partie inférieure de la distribution de l'output sur la base d'un critère de profit non-négatif. En effet, dans le contexte où il existe à la fois des économies d'échelle

et une situation de concurrence parfaite, les petits établissements ne pouvant couvrir leur coûts de production ne seront pas observés, ce qui causera un biais de sélection négatif aux estimations d'économies d'échelle provenant d'études économétriques. Une illustration de ce biais est présentée à la figure 3 où la vraie courbe de coût à long terme de l'industrie exhibe des rendements à l'échelle croissants et est représentée par la droite VV. De son côté, le prix unique du bien produit par cette industrie, p, délimite le seuil de survivance. Les firmes au-dessus de ce seuil ont un coût unitaire de production supérieur au prix et, par conséquent, quittent l'industrie à long terme. Ainsi, une analyse de régression faite à partir des données observées produira la fausse courbe de coût moyen FF qui sous-estimera la pente réelle de la fonction de coût.

Figure 3 : Illustration du problème de biais de sélection.



Quel est l'importance empirique de ce biais de sélection? En fait, l'importance empirique de ce dernier biais de sélection dépend largement des hypothèses d'économies d'échelle et de concurrence parfaite. Il existe une relation positive entre l'ampleur du biais d'une part, et le degré de concurrence et d'économies d'échelle d'autre part. Par exemple, si le niveau de concurrence est assez faible pour permettre l'existence d'un prix-parapluie permettant à toutes les firmes de survivre ou si les rendements à l'échelle sont constants, alors les estimations économétriques ne seront pas biaisées. Ainsi, ce problème d'estimation variera d'une industrie à l'autre selon le degré de concurrence et l'importance des économies d'échelle.[26]

5.4 L'hypothèse de prix des inputs constants

Plusieurs études excluent les prix des inputs de la partie déterministe de la fonction de coût moyen en supposant qu'ils sont constants d'une firme à l'autre à l'intérieur d'une industrie donnée. Or, si cette hypothèse n'est pas valide, la méthode utilisée dans ces études implique l'existence de biais de spécification causés par l'omission de variables explicatives. Sous l'hypothèse que le prix de l'input est corrélé positivement avec l'output, ceteris paribus, l'élasticité-coût sera biaisée positivement et l'élasticité d'échelle sera biaisée négativement. Puisqu'on observe généralement que le taux de salaire moyen augmente avec la taille des firmes (Miller (1981)), l'omission du taux de salaire implique probablement un biais négatif sur l'élasticité d'échelle[27]. Inversement, en supposant que le prix de l'input et l'output sont négativement corrélés, ceteris paribus, l'élasticité d'échelle sera biaisée positivement. Ce dernier scénario s'applique probablement à l'analyse de l'omission du prix de l'énergie[28].

---

26. Pour une présentation des méthodes d'estimation appropriées lorsque l'échantillon utilisé est tiré d'une distribution tronquée, on consultera Maddala (1983).

27. Cet argument est noté par Baldwin et Gorecki (1985) et Ware (1986).

28. Dans le cas où le prix de l'input et l'output ne sont pas corrélés, ceteris paribus, il n'existe pas de biais. Ces résultats découlent de l'analyse classique du signe d'un biais de spécification.

5.5 Mesure, diversité et endogénéité de l'output

Pour les analyses économétriques des structures de production, le traitement du niveau de production constitue sûrement un problème à plusieurs dimensions. Le premier problème et probablement le plus important résulte du fait que, même au niveau le plus désagrégé des classifications industrielles, le niveau de production observé correspond à la somme de plusieurs biens différents et que la part de chacun de ces biens dans l'output total varie d'un établissement à l'autre. L'hypothèse d'un bien composite homogène nécessaire à l'analyse est donc violée et l'impact sur les estimations d'économies d'échelle est très difficile à évaluer.

Cependant, le problème de diversité peut être surmonté par l'utilisation d'une fonction de production ou de coût multi-produits (Brown, Caves et Christensen (1985)). Malheureusement, cette approche ne peut s'appliquer si le nombre de produits est supérieur à trois, ce qui est le cas dans plusieurs industries.

Dans le cas où il y a plus de trois produits, l'inclusion d'un indice de diversité, tel que l'indice d'Herfindahl, dans la partie déterministe de la fonction de coût permet un contrôle, partiel cependant, de l'hétérogénéité de l'output. Traditionnellement, on a surtout insisté dans la littérature sur les déséconomies d'échelle provenant des problèmes de coordination lorsque le nombre de produits et de lignes de production augmentent. Par contre, récemment on a plutôt insisté dans la littérature sur les économies de diversité («economies of scope») résultant de l'augmentation du nombre de produits.[29] Il y a des économies de diversité s'il est moins coûteux de produire deux ou plusieurs produits dans un même établissement plutôt que dans deux ou plusieurs établissements distincts. Ainsi, même lorsque le niveau de production est tenu constant, l'effet net du nombre de biens produits

29. Le concept d'économie de diversité a été formalisé dans les travaux novateurs de W.J. Baumol, J.C. Panzar et R.D. Willis. Le lecteur intéressé consultera Panzar et Willis (1981) et Baumol, Panzar et Willis (1982). Par ailleurs, pour une application empirique on consultera Kim (1986).

sur les coûts unitaires est indéterminé. En effet, dans ce cas l'impact final d'une augmentation du nombre de biens produits correspond à la somme des effets d'économies (ou de déséconomies) de diversité et de déséconomies d'échelle liées à la réduction de la longueur moyenne des séries de production.

La mesure du niveau de production constitue une seconde difficulté de l'analyse. D'après Hazeldine (1984), dans un contexte de concurrence imparfaite, les coûts moyens plus faibles des firmes de grande dimension sont peut-être le reflet de prix de vente plus élevés plutôt que de coûts réels plus faibles lorsque l'output est mesuré en termes de valeur et non de quantité. Il s'agit là d'une question empirique que seul le calcul d'un prix d'output permettrait vraiment de vérifier. Notons cependant qu'il est difficile d'imaginer pourquoi les petits établissements ne profiteraient pas d'un prix plus élevé fixé par les établissements de grande dimension.

Troisièmement, l'approche duale assume que le prix des inputs et le niveau de production sont exogènes et que les firmes minimisent leurs coûts de production. Si le niveau de production est aussi une variable de choix pour les producteurs résultant d'un processus de maximisation des profits, alors les coefficients estimés seront potentiellement sujet à des biais de simultanéité.

Divers estimateurs de variable instrumentale (VI) peuvent être utilisés pour corriger ce problème de simultanéité. Dans les analyses de coupe transversale, un instrument naturel correspond à la valeur retardée de t période(s) du niveau de production. Cependant, selon les résultats de Griliches (1967), la valeur de l'output retardée d'une période constitue un trop bon instrument en ce sens qu'elle est trop fortement corrélée avec la valeur courante de l'output et que, par conséquent, les estimateurs des MCO et de VI produisent à toutes fins pratiques les mêmes résultats.[30] Il est alors difficile de faire

30. Notons cependant que Griliches (1967) utilise des donnés regroupées et non des données individuelles.

l'hypothèse d'absence de corrélation entre l'instrument et le terme d'erreur de la fonction de coût. De plus, les banques de données contenant de l'information longitudinale sont plutôt rares. Dans ce contexte d'information imparfaite, l'utilisation d'un instrument de rang à la Durbin (1954) constitue une alternative.[31] Par ailleurs, lorsque les prix de l'output et des inputs sont disponibles, l'utilisation d'un estimateur des moindres carrés à deux étapes est aussi possible. Il est cependant rare d'avoir accès à de tels prix dans les banques de données microéconomiques.

Notons finalement que l'utilisation d'un estimateur de variable instrumentale se fait au coût d'une perte sensible d'efficience. Par conséquent, il est préférable de tester l'existence de biais de simultanéité sur l'estimateur des MCO avant d'opter définitivement pour un estimateur de variable instrumentale.[32]

5.6 L'hypothèse d'homothéticité

Comme nous l'avons souligné plus haut, l'utilisation d'une fonction de coût «variable» peut créer un biais sur l'estimation des économies d'échelle si la fonction de production n'est pas homothétique. Plus précisément, les élasticités d'échelle estimées seront biaisées positivement (négativement) si les firmes de plus grande dimension ont une production relativement moins (plus) intensive en capital. Cependant, le peu d'évidence empirique disponible aux niveaux macroéconomique et microéconomique nous indique que l'hypothèse d'homothéticité ne biaise pas substantiellement l'estimation des économies d'échelle.

---

31. Dans le contexte de l'estimation d'une fonction de coût, il s'agit d'utiliser le rang de l'output dans la distribution de l'output comme instrument.

32. Les tests de spécification proposés par Durbin (1954) et Hausman (1978) sont des exemples de tests qui peuvent être utilisés. On consultera Nakamura et Nakamura (1981) pour une comparaison de ces tests.

5.7 Petits établissements: salaire et classification

Dans les études économétriques, l'exclusion des petits établissements constitue la règle plutôt que l'exception. Deux caractéristiques indésirables associées aux petits établissements expliquent ce traitement. D'abord, il existe une ambiguïté entre les revenus de travail et les profits dans les petits établissements. D'après Stigler (1963), le biais induit sur les coûts des petits établissements peut être à la fois positif ou négatif selon le type de comptabilité utilisée pour répartir le surplus d'opération entre les paiements au travail et au capital. Ensuite, puisque les établissements sont assignés à une industrie sur la base de leurs produits primaires, on s'accorde généralement pour dire que l'inclusion des petits établissements dans une industrie est arbitraire.

5.8 Hétéroscédasticité

À cause de la variance élevée du niveau de production, un problème d'hétéroscédaticité peut affecter l'estimation d'une fonction de coût. En présence de termes d'erreurs hétéroscédastiques, l'estimateur des MCO est non biaisé et convergent mais inefficient. De façon plus importante, l'estimateur-MCO de la matrice de variances-covariances des paramètres sera biaisé et non convergent.[33]

La solution classique à ce problème consiste à utiliser un estimateur des moindres carrés généralisés estimés (MCGE). Cependant, cet estimateur nécessite l'évaluation du modèle théorique engendrant les différences dans les variances des termes d'erreurs et la qualité des paramètres estimés et de leurs variances, en termes de convergence et

33. Plusieurs tests pour détecter la présence d'hétéroscédasticité ont été proposés dans la littérature économétrique (voir Johnston (1984, pp. 298-302)).

d'efficience, dépend fortement de la qualité du modèle déterministe utilisé. Conséquemment, l'estimateur des MCGE n'a pas nécessairement les propriétés de convergence et d'efficience.[34]

Lorsque la matrice de variances-covariances est inconnue, une alternative à l'estimateur des MCGE consiste à utiliser l'estimateur des MCO puisqu'il n'est pas biaisé, mais à utiliser une matrice de variances-covariances corrigée pour la présence de termes d'erreurs hétéroscédastiques. Cette matrice corrigée peut être obtenue à l'aide de la méthode proposée par White (1980). Ainsi, même si l'hétéroscédasticité ne peut être éliminée en termes de perte d'efficience, des inférences statistiques valides peuvent être effectuées. Le principal avantage de cette approche est qu'elle ne repose pas sur un modèle spécifique et possiblement incorrect de la structure d'hétéroscédasticité.

34. Évidemment, lorsque le processus qui engendre l'inégalité des variances est connu, l'estimateur des MCGE n'est pas affecté par ce problème. C'est le cas, par exemple, lorsqu'on utilise des données groupées et que l'hétéroscédasticité est causé par une inégalité du nombre d'observations d'un groupe à l'autre.

6. CONCLUSION

Au début des années soixante-dix, après un survol des études empiriques en économie de l'organisation industrielle, Leonard Weiss concluait que l'analyse empirique des économies d'échelle était «(a) fairly blank field» (1971, p. 397). Force est d'admettre qu'au milieu des années quatre-vingt cette conclusion tient toujours. En effet, depuis les années soixante-dix les efforts empiriques se sont surtout effectués dans quelques industries réglementées, comme le montre le survol de Jorgenson (1984). De leur côté, les études empiriques s'appliquant aux autres secteurs industriels sont encore trop peu nombreuses.

Ce qui frappe encore plus c'est l'absence d'évolution de la méthodologie utilisée par ces études durant toutes ces années. Si on exclut les études s'adressant aux secteurs réglementés, il faut en effet se rendre à l'évidence que la plupart des travaux microéconomiques récents continuent, à peu de chose près, à utiliser les mêmes techniques d'estimation que celles utilisées par Johnston (1960) qui fit figure de pionnier dans ce domaine. Cette absence d'évolution dans les études microéconomiques contraste avec le cheminement qu'ont suivi les études macroéconomiques où l'on a vu les formes fonctionnelles flexibles et l'estimation de système d'équations prendre de plus en plus de place au cours des années pour finalement devenir la «norme».

C'est avant tout du côté des données qu'il faut regarder pour expliquer cette quasi-stagnation des méthodes d'estimation utilisées dans les études microéconomiques depuis presque 30 ans. En effet, une amélioration de la précision et de la qualité des estimations d'économies d'échelle passe inexorablement par des efforts supplémentaires, à la base, au niveau des banques de données: la quantité, la qualité et la disponibilité des données microéconomiques sont autant de dimensions qu'il est nécessaire d'améliorer. Sur ce point, les organismes responsables de la collecte de ces données peuvent

avoir une contribution importante. Par exemple, la collecte auprès des établissements de données sur le capital et les coûts d'opération à l'intérieur d'une même enquête, et non séparement comme c'est souvent le cas, constituerait sûrement un pas important dans la bonne direction.

Par ailleurs, plusieurs changements peuvent être apportés aux analyses traditionnelles pour augmenter la précision des estimations à partir des banques de données existantes. Par exemple, la création de prix des facteurs de production et du niveau de production à l'aide de méthodes d'agrégation appropriées[35] permettrait de se rapprocher des concepts théoriques et d'évaluer l'importance de certains problèmes notés à la dernière section.

Finalement, certains peuvent s'interroger sur l'utilité d'estimations d'économies d'échelle dérivées d'études économétriques étant donné tous les problèmes et biais qui affectent ces estimations. Sur ce point, mentionnons que les comparaisons faites entre les études d'ingénierie et les études économétriques indiquent que les résultats économétriques sous-estiment substantiellement l'importance des économies d'échelle.[36] Ainsi, même si, de leur côté, les estimations d'ingénierie sont probablement biaisées à la hausse, il semble raisonnable de considérer les estimations économétriques d'économies d'échelle comme une borne minimale, du moins dans le cas du niveau minimal de production optimale. De plus, il existe généralement une forte corrélation entre les estimations des paramètres d'échelle tirés des analyses économétriques et d'ingénierie, de sorte que l'une ou l'autre de ces méthodes peut être utilisée pour évaluer les différences inter-industrielles d'économies d'échelle.

---

35. Sur ce point, on pourra consulter Robidoux (1986).

36. Par exemple, à partir d'un échantillon de 11 industries, Gupta et Fuss (1979) montrent que leurs estimations du niveau MES sont, en moyenne, 5 fois inférieures à celles produites par les analyses d'ingénierie. Par contre, il ne semble pas y avoir de biais négatif systématique au niveau de la pente de la fonction de coût.

Bibliographie

Baldwin, J.R., et P.K. Gorecki (1986), The Role of Scale in Canada-U.S. Productivity Differences in Manufacturing Sector: 1970-1979, Vol. 6, Collected Research Studies, Royal Commission on the Economic Union.

Baldwin, J.R., et P.K. Gorecki (1986a), «The Relationship Between Plant Scale and Product Diversity in Canadian Manufacturing Industries», The Journal of Industrial Economics, XXXIV, pp. 1-16.

Baumol, W.J., Panzar, J.C., et R.D. Willig (1982), Contestable Markets and the Theory of Industry Structure, New York; Harcourt Brace Javanovich.

Borts, G.H. (1960), «The Estimation of Rail Cost Functions», Econometrica, Vol. 28, No. 1, pp. 108-131.

Brown, R.S., Caves, D.W. et L.R. Christensen (1985), «Modelling the Structure of Cost and Production for Multiproduct Firms», Applied Economics, pp. 256-273.

Brown, R.S., et L.R. Christensen (1981), «Estimating Elasticities of Substitution in a Model of Partial Static Equilibrium: An Application to U.S. Agriculture, 1947 to 1974», in Berndt, E.R., et Field, B.C. ed., Modelling and Measuring Natural Resource Substitution, MIT Press, 1981, pp. 209-230.

Caves, D.W., Christensen, L.R., and J.A. Swanson (1981), «Productivity Growth, Scale Economies, and Capacity Utilization in U.S. Railroads, 1955-1974», American Economic Review, pp. 994-1002.

Comanor, W.S. et T.A. Wilson (1969), «Theory of the Firm and of Market Structures: Advertising and the Advantages of Size», American Economic Review, Papers and Proceedings, 59, pp. 87-98.

Corbo, V. et P. Meller (1979), «The Translog Production Function: Some Evidence from Establishment Data», Journal of Econometrics, Vol. 10, No. 2, pp. 193-199.

Cox, D. et R.G. Harris (1985), «Trade Liberalization and Industrial Organization: Some Estimates for Canada», Journal of Policital Economy, 93, 113-45.

Cox, D. et R.G. Harris (1986), «A Quantitative Assessment of the Economic Impact on Canada of Sectoral Free Trade with the United States», Canadian Journal of Economics, XIX, No. 3, pp. 377-394.

Daly, M.J., et P.S. Rao (1983), Productivity Growth, Economies of Scale, and Capacity Utilization in the Canadian Electric Power Industry: the Case of Ontario Hydro, Economic Council of Canada, Discussion Paper no. 236.

Daly, M.J., et P.S. Rao (1986), «Free Trade, Scale Economies and Productivity Growth in Canadian Manufacturing», The Manchester School, pp. 391-402.

De Brabander, B. et E. Vanlommel (1978), «Economies of Scale, Minimum Optimal Plant Size and Effectiveness of Market Structures in Belgian Industry anno 1970» European Economic Review, 11, pp. 363-377.

Delorme, F. et J. Lester (1986), The Structure of Production in Ten Canadian Industries, Working Paper No. 86-1, Fiscal Policy and Economic Analysis Branch, Department of Finance, Canada.

Denny, M., Fuss, M.A., et L. Waverman (1979), Energy and the Cost Structure of Canadian Manufacturing Industries, University of Toronto, Technical Paper no. 12.

Denny, M., et D. May (1977), «The Existence of a Real Value Added Function in the Canadian Manufacturing Sector», Journal of Econometrics, pp. 55-69.

Durbin, J. (1954), «Errors in Variables», Review of the International Statistical Institute, pp. 23-32.

Forsund, F.R. (1975), «The Homothetic Production Function», The Swedish Journal of Economics, Vol. 77, No. 2, pp. 234-244.

Friedman, M. (1955), «Comment on Statistical Cost Curves», In Business, Concentration and Price Policy, A Conference, Princeton University Press.

Fuss, M.A. (1977), «The Demand for Energy in Canadian Manufacturing: An Example of the Estimation of Production Structures with Many Inputs», Journal of Econometrics, 5, pp. 89-116.

Fuss, M.A., et V.K. Gupta (1981), «A Cost Function Approach to the Estimation of Minimum Efficient Scale, Returns to Scale, and Suboptimal Capacity», European Economic Review, 15, pp. 123-135.

Fuss, M.A., McFadden, D. et Y. Mundlak (1978), «A Survey of Functional Forms in the Economic Analysis of Production», dans Fuss, M.A. et McFadden, D., Production Economics: A Dual Approach to Theory and Applications, North-Holland, Amsterdam, pp. 219-268.

Fuss, M.A., et L. Waverman (1985), Productivity Growth in the Automobile Industry, 1970-1980: A Comparison of Canada, Japan and the United States, Working Paper No. 1735, National Bureau of Economic Research.

Gold, B. (1981), «Changing Perspectives on Size, Scale and Returns: An Interpretive Survey», Journal of Economic Literature, XIX, pp. 5-33.

Gorecki, P.K. (1976), Economies of Scale and Efficient Plant Size in Canadian Manufacturing Industries, Consumer and Corporate Affairs, Ottawa.

Green, C. (1985), Canadian Industrial Organization and Policy, McGraw-Hill, Second Edition.

Griliches, Z. (1967), «Production Functions in Manufacturing: Some Preliminary Results», in M. Brown, ed., The theory and Empirical Analysis of Production, New York: Columbia University Press, pp. 275-340.

Griliches, Z. et V. Ringstad (1971), Economies of Scale and the Form of the Production Function, Contribution to Economic Analysis No. 72, North-Holland, Amsterdam.

Gupta, V.K. (1968), «Cost Functions, Concentration, and Barriers to Entry in Twenty-nine Manufacturing Industries of India», The Journal of Industrial Economics, Vol. XVII, No. 1, pp. 57-72.

Gupta, V.K., M.A. Fuss (1979), Returs to Scale and Suboptimal capacity in Canadian Manufacturing: A Cost function approach, Working Paper No. 7904. University of Toronto.

Haldi, J., et D. Whitcomb (1967), Economies of Scale in Industrial Plants, Journal of Political Economy, 75, 4, pp. 373-385.

Hammond, C.J. (1986), «Estimating the Statistical Cost Curve: An Application of the Stochastic Frontier Technique», Applied Economics, 18, pp. 971-984.

Harris, R. (1984), «Applied General Equilibrium Analysis of Small Open Economies with Scale Economies and Imperfect Competition», American Economic Review, 74, 5, pp. 1016-1032.

Harris, R.G., with D. Cox (1984), Trade, Industrial Policy, and Canadian Manufacturing, Ontario Economic Council, Toronto.

Hausman, J.A. (1978), «Specification Tests in Econometrics», Econometrica, 46, 6, pp. 1251-1291.

Hazeldine, T. (1984), Trade, industrial policy and Canadian manufacturing: the Harris/Cox model and an alternative, Department of Agricultural Economics, University of British Columbia.

Johnston, J. (1958), «Statistical Cost Functions: A Re-appraisal», The Review of Economics and Statistics, Vol. XL, No. 4, pp. 339-350.

Johnston, J. (1960), Statistical Cost Analysis, McGraw-Hill.

Johnston, J. (1984), Econometric Methods, Troisième Edition, McGraw-Hill.

Jorgenson, D.W. (1984), Econometric Methods for Modelling Producer Behavior, Discussion Paper 1086, Harvard Institute for Economic Research.

Kim, H.Y. (1986), «Economies of Scale and Economies of Scope in Multiproduct Financial Institutions: Further Evidence from Credit Unions», Journal of Money, Credit and Banking, 18, 2, pp. 220-226.

Lyons, B. (1980), «A New Measure of Minimum Efficient Plant Size in UK Manufacturing Industry», Economica, 47, pp. 19-34.

Maddala, G.S. (1983), Limited-Dependent and Qualitative Variables in Econometrics, Cambridge University Press.

Miller, E.M. (1977), «Selection Bias in the Estimation of Cost and Production Functions», Journal of Economics and Business, 30, 1, pp. 79-81.

Miller, E.M. (1981), «Variation of Wage Rates with Size of Establishment», Economics Letters, pp. 281-286.

Nakamura, A., et M. Nakamura (1981), «On the Relationships Among Several Specification Error Tests Presented by Durbin, Wu, and Hausman», Econometrica, 49, 6, pp. 1583-1588.

Nerlove, M. (1963), «Returns to Scale in Electricity Supply» in C.F. Christ, ed., Measurement in Economics, Stanford, Standford University Press, pp. 167-198.

Panzar, J.C., et R.D. Willig (1981), «Economies of Scope», American Economic Review, 71, 2, pp. 268-272.

Rao, P.S., et R.S. Preston (1983), Inter-factor substitution and Total Factor Productivity Growth: Evidence from Canadian Industries, Economic Council of Canada, Discussion Paper no. 242.

Ringstad, V. (1974), «Some Empirical Evidence on the Decreasing Scale Elasticity», Econometrica, Vol. 42, No. 1, pp. 87-101.

Robidoux, B. (1986), Using the Divisia Aggregation Procedure for Creating an Establishment-Specific Energy Price Index, mimeo, Department of Finance, Canada.

Sato, K. (1975), Production Functions and Aggregation, Contribution to Economic Analysis No. 90, North-Holland.

Silberston, A. (1972), «Economies of Scale in Theory and Practice», The Economic Journal, Vol. 82, (March-Supplement), pp. 369-389.

Scherer, F.M. (1974), «Economies of Scale and Industrial Concentration», dans Goldschmid, H.S. et al. ed., Industrial Concentration: the New Learning, Columbia University for Law and Economic Studies and Little Brown.

Scherer, F.M. (1973), «The Determinants of Industrial Plant Sizes in Six Nations ,» Review of Economics and Statistics, pp. 135-145.

Scherer, F.M et al. (1975), The Economics of Multi-Plant Operations, Harvard University Press, Cambridge.

Scherer, F.M. (1980), Industrial market Structure and Economic Performance, Rand McNally, Second Edition.

Shaffer, S. (1985), «Competition, Economies of Scale, and Diversity of Firm Sizes,» Applied Economics, 17, pp. 467-476.

Stigler, G.J. (1963), Capital and Rates of Return in Manufacturing Industries, Princeton University Press.

Varian, H.R. (1978), Microeconomic Analysis, Norton.

Ware, R. (1986), Engineering Estimates of Scale Economies in United States Manufacturing Industry, University of Toronto, Mimeo.

Weiss, L. (1971), «Quantitative Studies of Industrial Organization», dans Intriligator, M.D., ed., Frontiers of Quantitative Economies, Contributions to Economic Analysis Series, Vol. 71, North Holland, pp. 362-403.

Whalley, J. (1984), ««Trade, Industrial Policy, and Canadian Manufacturing» by Richard G. Harris (with the assistance of David Cox): a review article,» Canadian Journal of Economics, pp. 386-198.

White, H. (1980), «A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity», Econometrica, 48, 4, pp. 817-838.

Zellner, A., et N.S. Revankar (1969), «Generalized Production Functions», Review of Economic Studies, 36, 241-250.

Zohar, V. (1982), Canadian Manufacturing: A Study in Productivity and Technological Change, Canadian Institute for Economic Policy.